Equity and efficiency in private and public education: a nonparametric comparison*

Laurens Cherchye † Kristof De Witte ‡
Erwin Ooghe §

Centre for Economic Studies, University of Leuven (KU Leuven),

Naamsestraat 69, 3000 Leuven, Belgium

Ides Nicaise ¶

Higher Institute of Labour Studies, University of Leuven (KU Leuven),

Parkstraat 47, 3000 Leuven, Belgium

October 9, 2007

Abstract

We present a nonparametric approach for the equity and efficiency evaluation of (private and public) primary schools in Flanders. First, we use a nonparametric (Data Envelopment Analysis) model that is specially tailored to assess educational efficiency at the pupil level. The model accounts for the fact that minimal prior structure is typically available for the behavior (objectives and feasibility set) under evaluation, and it reckons with outlier behavior in the available data, while it corrects for 'environmental' characteristics that are specific to each pupil. Second, we propose first- and second-order stochastic dominance (FSD and SSD) criteria as naturally complementary aggregation criteria for comparing the performance of different school types (private and public schools) in Flanders. While FSD only accounts for (Pareto) efficiency, SSD also takes (Pigou-Dalton) equality into consideration. We find that private schools outperform public schools in terms of SSD.

*We would like to thank Jan Van Damme for his permission to use the SiBO-data, Frederik Maes and Peter Helsen for their very useful help with the data, and the educational department of the Flemish government for the provision of some additional data.

†Laurens Cherchye is a post-doctoral Fellow of the Fund for Scientific Research - Flanders. E-mail to Laurens.Cherchye@kuleuven-kortrijk.be.

‡Email to Kristof.Dewitte@econ.kuleuven.be.

§Erwin Ooghe is a post-doctoral Fellow of the Fund for Scientific Research - Flanders. E-mail to Erwin.Ooghe@econ.kuleuven.be.

¶Email to Ides.Nicaise@hiva.kuleuven.be.

Keywords: equity; efficiency; private versus public education; non-parametric analysis; data envelopment analysis; stochastic dominance

1 Introduction

An important theme in policy evaluation is the question whether public funds are used in an equitable and efficient way. In the specific context of education, the comparison between private —but, possibly, publicly funded— schools and public schools is at the heart of a debate, which started with the work of Coleman *et al.* (1982). They find that (1) catholic school students obtain higher standardized test scores than public school students (after controlling for family background), and (2) catholic schools provide more equal educational outcomes for minority students. Therefore, one could conclude that catholic schools were both more efficient and more equitable than public schools in the U.S. at that time. The work of Coleman *et al.* was (and still is) controversial, not only in the public debate (see, e.g., the New York Times articles of April 7, April 12 and April 26, 1981, discussing the consequences of Coleman *et al.*'s results for the introduction of tuition tax credits and/or school vouchers), but also in academics (see, e.g., Cain and Goldberger (1983) for a neat overview of methodological problems). In spite of these criticisms, many studies have confirmed the outperformance of public by private schools; see, e.g., the literature review in Altonji *et al.* (2005).

This study compares private and public primary schools in Flanders, i.e., a region in Belgium, on the basis of both equity and efficiency considerations. Our methodology consists of two steps, a measurement and an aggregation step. The distinguishing feature is that both steps are entirely nonparametric. First, we use a nonparametric efficiency evaluation model —also called a Data Envelopment Analysis (DEA) model— which is specially tailored for environment-corrected educational efficiency evaluation at the pupil level. Second, we use nonparametric stochastic dominance techniques —which allow us to take efficiency and equity considerations into account— to compare the aggregate performance of private and public schools. While our focus is on comparing school types, this aggregation step could alternatively be implemented e.g. for performance assessments at the school level. In addition, although our application focuses on education, we believe that the suggested method is applicable in a wide variety of public sector settings (e.g., health services), which typically involve not only efficiency but also equity considerations.

To set the stage, we briefly present the measurement and aggregation step in more detail and relate them to the existing literature. We use a nonparametric DEA model to measure educational efficiency at the pupil level on the basis of test scores in mathematics and language proficiency (writing and reading in Dutch). We account for the inputs used (which the policy makers do control) as well as for possibly diverging 'environmental' variables —socio-economic status of parents and lagged test score results— that might affect pupil performance (and which often fall beyond the control of policy makers and schools). DEA has the attractive feature that it imposes minimal *a priori* structure on the behavior (objectives and feasibility set) that is evaluated. This is particularly convenient in the context of primary education, where little *a priori* information is available; as such, the use of DEA minimizes the risk of specification error.

DEA models have been used before to evaluate the educational efficiency at the pupil level; see, e.g., Grosskopf *et al.* (1997, 1999) and Portela and Thanassoulis (2001) and the references therein. In the current study, we propose a DEA model that is specially designed for educational efficiency evaluation: while at the input side it uses the minimal 'free disposability' assumption (*in casu*, more input never leads to a lower (potential) performance), at the output side it uses the linear aggregation that is typical for measuring pupil performance in primary education (i.e., aggregate performance results are conventionally defined as weighted sums of the results in separate disciplines). Focusing on linearly aggregated output, it measures educational inefficiency in terms of the difference between the maximally attainable output and the actually achieved output.

Two additional features of our DEA model are worth mentioning. First, it uses linear output aggregation, but it allows for flexible weights for the different performance dimensions. Essentially, such a flexible weighting allows each pupil to be evaluated in terms of his/her own 'most favorable' weighting scheme, which accounts for 'specialization' in education; we avoid undesirable 'extreme' specialization by limiting the range of possible output weights through pre-specified bounds. Second, by suitably adapting the methodology of Daraio and Simar (2005, 2007) to our DEA model, it can account for outlier behavior, while it also allows us to explain observed performance differences in terms of diverging environmental characteristics in a nonparametric way. The observed environmental impact as well as the corresponding environment-corrected efficiency results provide an easy-to-implement tool for attention-direction in the political process.

Finally, we suggest first-order and second-order stochastic dominance criteria (also known as, respectively, 'rank dominance' and 'generalized Lorenz dominance' in the normative welfare literature) for comparing the aggregate performance of public and private schools; see, e.g., Lambert (2001) and Levy (1992) for surveys of stochastic dominance criteria in the welfare and risk literature, respectively. These criteria allow us to compare the social welfare loss in public and private schools, i.e., the difference between the maximally attainable welfare and the actual welfare of their pupils. We believe these criteria are particularly useful in the context of DEA efficiency evaluation of the public sector. First, they are nonparametric in nature, which naturally complies with the nonparametric orientation of DEA. Next, the second order stochastic dominance criterion considers not only aggregate (Pareto) efficiency but also expresses a concern for inequality, which is particularly relevant within the context of public policy evaluation. As with DEA, these aggregation criteria are easy-to-implement, which makes them attractive for practical applications.

The remainder of this paper unfolds as follows. The next section presents our research question. Section 3 discusses our methodology for evaluating educational efficiency at the individual pupil level. Section 4 presents the efficiency results, with a main focus on environmental effects. Section 5 discusses the aggregation of the individual efficiencies. A final section 6 summarizes our main conclusions.

2 Motivation

The general belief is that private (mainly catholic) schools in Flanders perform better (i.e., the cognitive output of their pupils is thought to be higher on average), but this statement is somewhat blurred by two counteracting forces related to inputs and environment. While private schools are said to have more pupils with an 'advantageous' family background, they must also receive less funding as a consequence of the 'Equal Educational Opportunities' programme of the Flemish government. In this section, we will define and describe the inputs, outputs and environment in the Flemish educational system.

We use data from the SiBO-project, whose aim is to describe and explain differences in the primary school curriculum of a cohort of Flemish pupils. The dataset consists of a reference group, which is representative for the Flemish population of primary school pupils, and three additional data sets: (1) all public city schools of the city of Ghent, (2) an oversampling to get a sufficient number of schools with a high number of disadvantage pupils (pupils for whom the schools get additional means in the so-called 'Equal Educational Opportunities' programme of the Flemish government) and (3) an oversampling to obtain a sufficient number of non-traditional schools. We use all pupils together, while correcting for the sample's non-representative nature in our empirical efficiency evaluation. This leaves us with 3413 pupils (with complete data), of whom 1774 attend private catholic schools, 1039 local public schools and 553 Flemish public schools. The remaining 47 pupils take classes in private non-catholic schools. Although these pupils are taken into account to estimate inefficiency scores later on, we use the term private to refer to pupils in catholic private schools in the sequel.

We look at the cohort of pupils in their second year of primary education (2004-2005) —at the (normal) age of 7— while we use data from the same pupils in the first year (2003-2004) to retrieve environmental variables. We extract 3 types of variables at the individual level, called inputs, outputs and environmental variables in the sequel.

Financial inputs in primary schools mainly consist of salaries (80%) and operation costs (20%). As we *a priori* assume that the differences in operation costs are unlikely to cause differences in cognitive results, we only focus on inputs related to teaching. Government assigns instruction units to pupils, which can be used by their respective schools to finance teachers: 24 instruction units correspond with a full-time teacher. The total number of instruction units assigned to a particular pupil consists of regular and additional, so-called 'equal educational opportunity' (EEO), instruction units. Regular (per-capita) instruction units (REG) are, roughly speaking, the same for all pupils, as they are divided among schools on the basis of a scale which is approximately linear in the number of pupils. The additional *EEO* instruction units depend on certain 'disadvantageous' pupil characteristics, to wit, the household income consists of replacement incomes only, the pupil is living outside the biological family, the level of education of the mother is low, the pupil's family belongs to a travelling population and —in combination with one of the former characteristics— the home language is different from Dutch. Table 1 contains some summary statistics for both types of instruction units *REG* and *EEO* over the different school types in Flanders.[1] Overall, local public schools receive most instruction units (per capita), private schools the least, while the Flemish public schools are in between both.

Table 1: (Input) *REG* and *EEO* instruction units per school type.

	school type	all	private	public	
input				local	Flemish
all	average	1.00	0.97	1.07	1.04
	std. dev.	0.28	0.26	0.30	0.28
REG	average	0.88	0.87	0.92	0.86
	std. dev.	0.18	0.18	0.19	0.15
EEO	average	0.12	0.09	0.15	0.19
	std. dev.	0.19	0.17	0.20	0.22

Output is defined on the basis of test scores in three dimensions: mathematics, technical reading and writing, collected at the end of the second year. All scores are set between 0 and 100. We calculate a language proficiency score as the simple average of the reading and writing scores. Table 2 provides summary statistics for the mathematics (*MATH*) and language proficiency score (*DUTCH*) for the different school types in Flanders. Private (catholic) schools do best in both tests. They are followed closely by the local public schools and, at some distance, by the Flemish public schools. Note also that the dispersion in test scores in the private (catholic) schools is smaller compared to local public schools, and dispersion in the latter type of schools is in turn smaller compared to Flemish public schools.

Table 2: (Output) *MATH* and *DUTCH* per school type.

	school type	all	private	public	
output				local	Flemish
MATH	average	57.08	58.33	57.54	50.74
	std. dev.	19.40	18.95	19.02	20.78
DUTCH	average	55.27	56.49	54.00	51.56
	std. dev.	14.05	13.46	14.19	15.71

[1] All reported figures in this paper are weighted by the inverse of the sampling probability, to correct for the non-representative nature of the dataset.

Pupil environment is measured by three indices: socio-economic status and entry level in mathematics and language proficiency. Socio-economic status (*SES*) reflects the cultural, social and economic environment of the pupil's home. It is calculated as the average of the following three variables (after standardization): average education level, average professional status and total income of the parents of the pupil; see Reynders *et al.* (2005) for details. The begin level in mathematics (*B-MATH*) and language proficiency in Dutch (*B-DUTCH*) reflect the intellectual antecedents of the pupil, and is equal to the mathematics and language proficiency score of the pupil at the end of the previous year.

Table 3 reports summary statistics for *SES*, *B-MATH* and *B-DUTCH*. We find that, on average, private (catholic) schools attract pupils with more 'advantageous' environmental characteristics compared to local public schools and —to an even greater extent— Flemish public schools. Notice that the differences in *EEO* instruction units between the different school types (reported in Table 1) reflect the differences in *SES*.

Table 3: (Environment) *SES*, *B-MATH* and *B-DUTCH* per school type.

	school type	all	private	public	
environment				local	Flemish
SES	average	0.03	0.11	-0.01	-0.35
	std. dev.	0.85	0.83	0.85	0.83
B-MATH	average	53.26	53.91	54.05	48.44
	std. dev.	19.06	18.66	18.75	20.69
B-DUTCH	average	47.25	47.92	47.29	43.92
	std. dev.	9.74	9.69	9.42	9.87

To summarize, our data roughly confirm the widely held belief that private (catholic) schools in Flanders perform better, while they receive less teaching inputs as a consequence of their more 'advantageous' pupil population. Our main research question is how we must assess these output differences in a fair way, i.e., by taking the differences in inputs and environment into account.

3 Efficiency measurement: method

Consider a general educational system that is characterized, at the level of the pupil, by p inputs and q outputs. We denote the corresponding input vector by $x \in \mathbb{R}_+^p$, and the output vector by $y \in \mathbb{R}_+^q$; in our application, $p = 1$ and the input is the sum of the *REG* and *EEO* instruction units, while $q = 2$ and the outputs are the *MATH* and *DUTCH* scores. The set of all feasible combinations of educational inputs and outputs is the *feasibility set*

$$F = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\right\}.$$

Educational efficiency analysis relates educational input to educational output. As such, empirical efficiency evaluation essentially requires two steps: (1) we need to empirically estimate the feasibility set F; (2) we have to evaluate observed efficiency by using an efficiency measure that has a meaningful interpretation in terms of the underlying educational objectives. These two issues are discussed next. Subsequently, we discuss two additional issues that will be important for our empirical application: (3) we need to account for outlier observations in the empirical efficiency evaluation; and (4) we want to correct the observed efficiency scores for environmental characteristics, which will also allow us to visualize the impact of the latter on the former.

3.1 Empirical feasibility set

Usually, the 'true' feasibility set F is not observed. To deal with such incomplete information, the nonparametric approach suggests to start from the set of n observed input-output vectors $S \subseteq F$ ($|S| = n$); it assumes that observed input-output combinations are certainly feasible (e.g., Varian, 1984). In addition, we assume that inputs and outputs are freely disposable, which means:

$$\text{if } (x, y) \in F \text{ then } (x', y') \in F \text{ for } x' \geq x \text{ and } y' \leq y.$$

Taken together, these assumptions lead to the *empirical feasibility set*

$$\widehat{F} = \left\{ (x, y) \in \mathbb{R}_{+}^{p+q} \mid \ x' \geq x \text{ and } y' \leq y \text{ for } (x', y') \in S \right\};$$

i.e., the free disposal hull (FDH) of the set S (e.g., Deprins *et al.*, 1984; Tulkens, 1993).

We briefly discuss the interpretation of the assumptions that underlie the construction of $\widehat{F}$. First, 'free disposability of inputs' means that more input never implies a decrease of the (maximally achievable) output. We believe this is a reasonable assumption in the current context, where inputs stand for instruction units and outputs stand for pupil performance (in alternative disciplines). Second, 'free disposability of outputs' means that more output never implies a decrease of the (minimally required) input. Once more, we believe this assumption is tenable in our specific context.

Finally, the assumption $S \subseteq F$ excludes measurement errors and atypical observations, such that all observed input-output vectors are comparable (or, alternatively, that all relevant input and output dimensions are included in the analysis). Admittedly, this assumption may seem problematic in our application, which compares primary pupils that may be characterized by different background characteristics (that are not explicitly included in our set of conditioning/environmental variables; see further: conditional inefficiency measure). Therefore, as we will explain further on, we will use an efficiency evaluation method that mitigates the impact of potential outliers within the observed set S.

3.2 Inefficiency measure

Consistent with the usual practice in primary education, we focus on output performance (see, e.g., Worthington, 2001). Specifically, we use an inefficiency measure which is, for a given input, equal to the maximally possible output performance minus the actual output performance. The output performance is measured as a weighted sum of the output performances in alternative disciplines (captured by the q constituent components of each output vector y), which again reflects the usual practice in primary education. Suppose, that we are to evaluate a pupil observation $(x_E, y_E) \in S$ (also referred to as 'observation E' in what follows) and that the relevant output weights are given by $w_E \in \mathbb{R}^q_+$. For the empirical feasibility set $\widehat{F}$, educational inefficiency for this pupil is defined as

$$\theta_E = \max_{(x,y)\in\widehat{F}} \left\{ \frac{w_E \cdot (y - y_E)}{w_E \cdot g} \mid x \leqslant x_E \right\},$$

with $g \in \mathbb{R}^q_+$ an aggregation vector that defines the denominator as a weighted sum of the output weights; we use $w_E \cdot g > 0$. For the given input level, the measure takes the difference of (linearly aggregated) maximal output performance over actual output performance; this difference is normalized by dividing through the weighted sum $w_E \cdot g$. Clearly, $\infty > \theta_E \geq 0$. Efficiency implies $\theta_E = 0$; and higher inefficiency values generally reveals more inefficiency. In our application, we set the aggregation vector g equal to a q-dimensional vector of ones, which implies that the denominator is simply the (equally weighted) sum of weights. We believe this specification of g is appropriate in our application context because the outputs ($MATH$ and $DUTCH$) are measured in a comparable measurement unit: it naturally corrects for the scale of the output weights w_E (i.e. κw_E obtains the same results as w_E for all $\kappa > 0$), while treating the (directly comparable) output dimensions identically. But it should be clear that, in general, our method also allows for other specifications of g, which accounts for the possibility that different outputs are expressed in different measurement units.[2]

The measure θ_E assumes that the weighting vector w_E is fixed *a priori*. Our following application will focus on an alternative inefficiency measure that allows for flexible weighting. Specifically, for each pupil observation we choose 'most favorable' weights $\widehat{w}_E$ that maximize the efficiency of the input-output vector under evaluation; this conveniently allows for 'specialization' in learning: e.g. if a pupil performs relatively well in mathematics, then this discipline gets a relatively high weight in her/his inefficiency measure. To avoid undesirable 'extreme' specialization, we impose that the endogenously selected relative output weights $\widehat{w}_E$ should respect upper and lower bounds, which are captured by the set $W_E \subseteq \mathbb{R}^q_+$ characterized in terms of linear constraints ($\widehat{w}_E \in W_E$ satisfying $\widehat{w}_E \cdot g > 0$). (The construction

[2] In this respect, it is also worth indicating that, for general g, the 'empirical' inefficiency measure $\widehat{\theta}_E$ (cfr. infra) is formally similar to the so-called 'directional distance function'; see, for example, the duality results in Chambers *et al.* (1998, p. 358). These authors also provide a discussion on possible specifications of g; while they focus on profit efficiency, the analogy with our setting is straightforward.

of W_E for our empirical application is discussed in the beginning of section 4.) This yields the empirical *inefficiency measure*

$$\widehat{\theta}_E = \min_{\widehat{w}_E \in W_E} \max_{(x,y)\in \widehat{F}} \left\{ \frac{\widehat{w}_E \cdot (y - y_E)}{\widehat{w}_E \cdot g} \mid x \leqslant x_E \right\}.$$

Clearly, for $w_E \in W_E$ we have $\theta_E \geq \widehat{\theta}_E \geq 0$. The measure $\widehat{\theta}_E$, with endogenously defined most favorable weights, has a directly similar interpretation as the measure θ_E, with *a priori* fixed weights w_E.

To conclude, we note that the empirical inefficiency measure can be computed by simple linear programming. Specifically, given the construction of $\widehat{F}$, the computation proceeds in two steps. The first step identifies the set of observations that dominate the evaluated observation in input terms:

$$D_E = \{(x, y) \in S \mid \ x \leqslant x_E\}.$$

The second step involves the linear programming problem. As a preliminary note, we recall that $\widehat{w}_E \cdot g > 0$ in the above definition of $\widehat{\theta}_E$, so that we can use the normalization $\widehat{w}_E \cdot g = 1$ (because the set W_E only restricts the relative output weights). As such, we can compute

$$\widehat{\theta}_E = \min_{u,\widehat{w}_E \in W_E} \left\{ u - \widehat{w}_E \cdot y_E \mid \begin{array}{c} \widehat{w}_E \cdot g = 1 \\ u \geq \widehat{w}_E \cdot y \quad \forall y : (x, y) \in D_E \\ \widehat{w}_E \in W_E \end{array} \right\}.$$

This is a linear programming problem given that the set W_E is characterized by linear constraints. For general W_E, the fact that merely linear programming is required for the computation of the empirical inefficiency measure $\widehat{\theta}_E$ (after a trivial check of input dominance) makes it attractive for practical applications.

3.3 Outlier-robust inefficiency measure

To mitigate the impact of (potential) outlier behavior in the observed sample S, we use the order-m method as suggested by Cazals *et al.* (2002); we adapt the method for the specific inefficiency measure $\widehat{\theta}_E$ defined above. Essentially, in terms of the terminology introduced above, this boils down to repeatedly drawing (with replacement) R subsamples $D_E^{r,m}$ ($r = 1$, ..., R) from the dominating set D_E; each subsample $D_E^{r,m}$ contains m (> 1) input-output vectors that are selected from D_E, i.e. $D_E^{r,m} \subseteq D_E$ and $|D_E^r| = m$.[3] For each $D_E^{r,m}$ we

[3]Remark that, to correct for the non-representative nature of our dataset, we take the probability of drawing a pupil proportional to the inverse of the probability that this pupil appears in the sample due to the specific sampling design. A similar qualification applies to the environment-corrected inefficiency measure where we weight the Kernel functions by the inverse of the sampling probability.

compute the corresponding empirical inefficiency measure

$$\widetilde{\theta}_E^{r,m} = \min_{u,\widehat{w}_E \in W_E} \left\{ u - \widehat{w}_E \cdot y_E \;\middle|\; \begin{array}{l} \widehat{w}_E \cdot g = 1 \\ u \geq \widehat{w}_E \cdot y \quad \forall y : (x,y) \in D_E^{r,m} \\ \widehat{w}_E \in W_E \end{array} \right\},$$

which again uses linear programming. Subsequently, the outlier-robust *order-m inefficiency measure* is defined as the arithmetic average

$$\widetilde{\theta}_E^{m} = \frac{\sum_{r=1}^{R} \widetilde{\theta}_E^{r,m}}{R}.$$

Referring to Cazals *et al.* (2002), this measure has attractive statistical properties and conveniently mitigates outlier behavior. See also Simar (2003) for a related discussion.[4] As a final note, because it can well be that $(x_E, y_E) \notin D_E^{r,m}$, we can have $\widetilde{\theta}_E^{r,m} < 0$. We will label such observation as 'super-efficient' in what follows.

3.4 Environment-corrected inefficiency measure

To capture environmental effects, we use the procedure outlined by Daraio and Simar (2005, 2007). Like before, we adapt this method to the specific inefficiency measure under consideration.

Suppose we want to take up k environmental characteristics, which corresponds to a k-dimensional vector z of environmental indicators associated with each input-output vector (x, y); in our application, $k = 3$ and the vector z captures *SES*, *B-MATH* and *B-DUTCH*. For the evaluated observation E, the Daraio-Simar procedure computes an environment-corrected inefficiency measure by conditioning on the corresponding value z_E of the environmental vector: it selects input-output vectors $(x, y) \in D_E$ with z in the neighborhood of z_E. This gives us the *conditional inefficiency measure*

$$\widehat{\theta}_E(z_E) = \min_{u,\widehat{w}_E \in W_E} \left\{ u - \widehat{w}_E \cdot y_E \;\middle|\; \begin{array}{l} \widehat{w}_E \cdot g = 1 \\ u \geq \widehat{w}_E \cdot y \quad \forall y : (x,y) \in D_E(z_E) \\ \widehat{w}_E \in W_E \end{array} \right\},$$

with $D_E(z_E) = \{(x, y) \in D_E \mid |z_E - z| \leq h\}$ and h a Kernel bandwidth vector. In our application, when the number of conditioning variables k is larger than 1, we first apply a so-called Mahalanobis transformation to decorrelate the environmental variables (see, e.g., Mardia *et al.*, 1979). Afterwards, we perform a sequential Kernel estimation —as if all environmental variables were independently distributed— to compute the optimal bandwidth vector (via

[4] Cazals *et al.* (2002) actually consider an efficiency measure that does not consider linear but monotonic aggregation of the outputs. But their main results carry over to the linear variant that we consider. A similar qualification applies for our use of the procedure of Daraio and Simar (2005) to account for environmental effects in the efficiency evaluation exercise. In fact, these authors also focus on input efficiency, while we translate their procedure towards output efficiency.

the likelihood cross-validation criterion) and the probability weights used to draw the sample of size m.

4 Efficiency measurement: application

In this section, we focus on visualizing the impact of the environmental variables *SES*, *B-MATH* and *B-DUTCH* on educational efficiency at the pupil level, by using the outlier-robust order-m inefficiency measures described in the previous section. For these measures, an additional consideration concerns the specification of the paramaters R (the number of drawings with replacement) and m (the number of input-output vectors selected from D_E in each drawing). In the following, we discuss empirical results for $R = 50$ and $m = 100$ as, from these values on, the number of super-efficient observations (see supra) in the sample is robust at around 1%; the same criteria is used by Daraio and Simar (2007). Still, at this point it is worth stressing that we have also experimented with other values for R ($R = 10, 25, 100$) and m ($m = 10, 25, 50, 125, 150$); these alternative configurations generally obtained the same qualitative conclusions. For compactness, we do not include all these results in the current paper, but they are available from the authors upon simple request.

As discussed before, our application avoids 'extreme' specialization in either *DUTCH* or *MATH* by focusing on a restricted set $W_E \subseteq \mathbb{R}^q_+$ (with $q = 2$), which captures upper and lower bounds of the relative output weights. To construct these bounds, we divide the number of hours spent on *DUTCH* in the classroom by the sum of the number of instruction hours spent on *DUTCH* and *MATH*. This reflects the weight attached to *DUTCH* (relative to *MATH*) in the second year of primary education. The average equals 0.54 —and is very similar for the different school types— while the 1 and 99-percentile values equal 0.44 and 0.71, respectively. These 1 and 99-percentile values will serve as (relative) weight restrictions for *DUTCH* (and hence 0.56 and 0.29 for *MATH*). To check sensitivity of our main results with respect to this particular specification of W_E, we have also considered extreme scenarios with no weight flexibility (i.e. using 0.50 as a fixed weight for the two outputs *DUTCH* and *MATH*) and full weight flexibility (i.e. $W_E = \mathbb{R}^q_+$, with $\widehat{w}_E \cdot g = 1$ for $\widehat{w}_E \in W_E$). Our main qualitative results appeared to be robust for these alternative weight bounds; the corresponding results will not be reported in the current paper, but they are available from the authors upon simple request.

4.1 Outlier-robust inefficiency measures

Before visualizing the impact of the different environmental variables under study, Table 4 provides summary statistics for alternative outlier-robust order-m inefficiency measures. We report results for the full sample (see the column 'all') and for the subsamples that correspond to the different school types (private schools, local public schools and Flemish public schools).

Table 4: Some summary statistics for the robust inefficiency measures.

environment	school type	all	private	public local	Flemish
$\emptyset$	average	26.99	25.74	27.68	31.61
	std. dev.	13.85	13.28	13.78	15.49
	minimum	-5.70	-3.50	-3.67	-5.70
	maximum	74.10	74.03	74.10	71.39
SES	average	26.97	25.39	27.43	31.17
	std. dev.	13.80	13.25	13.78	15.27
	minimum	-3.02	-2.85	-3.02	-2.34
	maximum	75.96	73.45	75.96	70.03
B-MATH	average	24.17	23.00	25.12	27.88
	std. dev.	12.34	11.87	12.28	13.62
	minimum	-5.74	-5.42	-4.92	-5.74
	maximum	72.46	61.63	71.52	72.46
B-DUTCH	average	23.61	22.61	24.34	27.04
	std. dev.	12.35	11.80	12.39	14.03
	minimum	-6.54	-1.37	-6.54	-0.86
	maximum	65.41	62.88	65.41	60.09
B-MATH, B-DUTCH & SES	average	17.18	16.34	18.52	18.70
	std. dev.	10.16	9.78	10.35	11.04
	minimum	-17.52	-1.99	-17.52	-3.39
	maximum	55.14	49.72	55.14	53.43

Let us first regard the unconditional inefficiency values (with environment = $\emptyset$). Table 4 reports an average inefficiency score of 26.99 over all pupils in our sample. In words, the average pupil achieves an output level that is 26.99 points below the best possible performance for (at most) the same amount of instruction units (= $REG + EEO$ = input). To interpret this result, we recall that aggregate output performance is measured as a weighted sum of the output performance in the disciplines *MATH* and *DUTCH* (using 'most favorable' weights for each individual pupil), and that the *MATH* and *DUTCH* scores are both set between 0 and 100. As such, this average shortage of 26.99 points should be compared to a ('theoretical') maximum possible shortage of 100 points. Next, we also observe much variation in the efficiency scores over pupils. For example, the standard deviation in the inefficiency values is 13.85; and the maximum inefficiency value amounts to 74.10 points, while the minimum value equals -5.70.[5] Finally, we find differences in the distributions for different school types; for example, the average inefficiency value for private schools (25.74)

[5] We recall that negative inefficiency values are possible for super-efficient observations because we focus on outlier-robust inefficiency measures.

is below that for local public schools (27.68), which in turn is below that for Flemish public schools (31.61).

In the following, we investigate to what extent these patterns in the distribution of the inefficiency scores can be attributed to environmental differences, as captured by the variables *SES*, *B-MATH* and *B-DUTCH*. The summary statistics in Table 4 provide some preliminary insights. We first consider the separate impact of the social and cultural environment of a pupil's home (captured by *SES*) and the cognitive antecedents of the pupil (captured by *B-MATH* and *B-DUTCH*). As expected, we find that all three variables influence the pupils' efficiency values; for example, when focusing on the full sample (see the column 'all'), average inefficiency reduces to 26.97, 24.17 and 23.61 when controlling for, respectively, *SES*, *B-MATH*, and *B-DUTCH*. In addition, we observe a decrease in the variation of the inefficiency values; for example, the standard deviation reduces to 13.80, 12.34 and 12.35 when conditioning on, respectively, *SES*, *B-MATH* and *B-DUTCH*. This indicates that each individual variable can explain the observed variation in the inefficiency values to some extent. Finally, if we simultaneously control for *SES*, *B-MATH* and *B-DUTCH*, we observe a further and rather substantial decrease of the average inefficiency value (to 17.18 for 'all') as well as the standard deviation of inefficiency values (to 10.16 for 'all'). This suggests that simultaneous consideration of all three environmental variables can effectively yield additional 'explanatory' value in terms of explaining patterns of educational inefficiency. The same general conclusions hold for all three school types (private schools, local public schools and Flemish public schools). Remark, finally, that for all specifications of the conditioning variables that we consider, private schools are, on average, more efficient than both types of public schools, and that local public schools outperform Flemish public schools.

4.2 Environmental effects

To visualize environmental effects and, consequently, to detect whether an environmental variable is favorable or unfavorable, we adapt Daraio and Simar (2007)'s methodology to our setting. If z_E^{-j} denotes the vector of all conditioning variables, except for the j-th entry, and z_E^j is the j-th entry, then we can nonparametrically regress the differences $\widetilde{\theta}_E^m(z_E) - \widetilde{\theta}_E^m\left(z_E^{-j}\right)$ on the observed values for z_E^j. If, for a certain range, the regression is decreasing, the j-th environmental variable is unfavorable to efficiency, behaving as a 'substitutive' output in the educational process. Conversely, an increasing curve indicates a favorable variable that plays the role of a 'substitutive' input in the educational process. Finally, a flat curve suggests that there is no efficiency effect of the environmental variable.

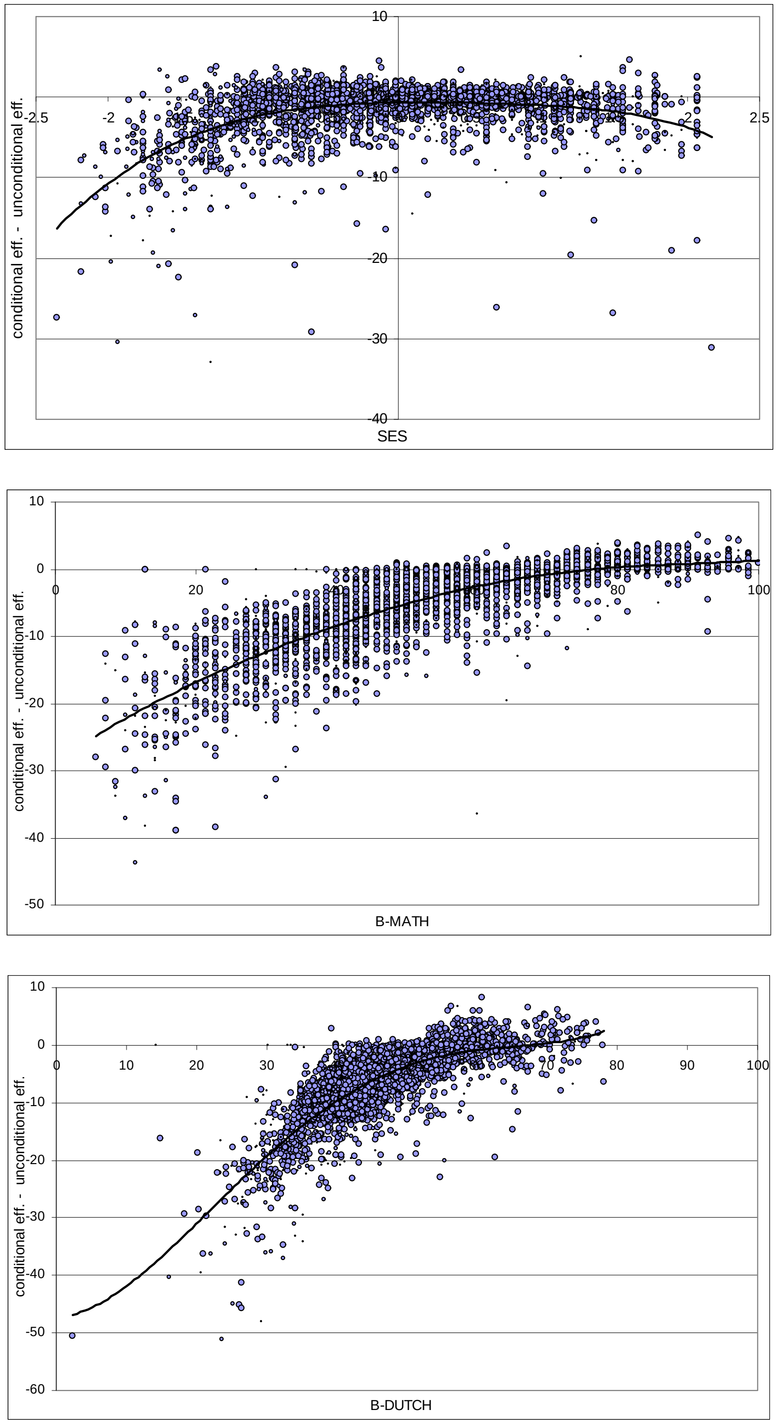

Figure 1: (Environmental impact) *SES*, *B-MATH* and *B-DUTCH.*

Figure 1 visualizes the environmental effects. We first consider the variable *SES.* Generally, we find a positive first order effect of *SES* on the educational efficiency for low *SES* values, and a negative first order effect for high *SES* values. The full line suggests a negative second order effect. Still, the effect for high *SES* values is not very pronounced; in fact, the observation points are widely scattered around the full line. We infer that, while *SES* admittedly has some (positive) effect on educational efficiency, much of this effect is already captured by the other two variables *B-MATH* and *B-DUTCH*, which causes the residual impact of *SES* to be rather low.

Let us then regard the variable *B-MATH*. Figure 1 reveals a positive monotonic first order impact and a generally negative second order effect. Taken together, this indicates that, on average, a higher *B-MATH* value predicts a higher educational efficiency, but the marginal impact decreases when the *B-MATH* value increases. Compared to the *SES* picture, the observation points are much more narrowly scattered around the full line, which provides more convincing support for this residual *B-MATH* effect.

Finally, we consider the variable *B-DUTCH*. The general conclusions drawn from Figure 1 are similar to those for *B-MATH*: there is a clearly positive monotonic first order effect; and, generally, a negative second order effect, which is now even more pronounced than in the *B-MATH* case. As before, we infer that, on average, a higher *B-DUTCH* value leads to a higher educational efficiency, but the marginal impact decreases when the *B-DUTCH* score increases (*in casu*, at a relatively rapid rate). The fact that the observation points are narrowly scattered around the full line implies quite strong support for this conclusion.

The overall conclusion, which falls in line with our prior expectations, is that each of the environmental variables positively impacts the educational efficiency (see the first order effects), and that this positive effect prevails in particular for low initial values for *SES*, *B-MATH* and *B-DUTCH* (see the second order effects). Although we find stronger effects for *B-MATH* and *B-DUTCH* than for *SES*, we believe that our results provide sufficiently strong support for simultaneously conditioning on all three variables when comparing the educational efficiency for different pupils. Therefore, our aggregation exercise in the next section will mainly focus on such fully conditioned educational efficiency values.

5 Aggregation: efficiency versus equity

This section aims to compare the aggregate efficiency and equity performance of private schools, local public schools and Flemish public schools. Specifically, we start with the pupils' inefficiency values and the corresponding optimal weights that underlie the results presented in the previous section. Using these pupil-specific weights to aggregate *DUTCH* and *MATH*, we obtain what we will call the 'actual score'. Adding the inefficiency score to it, we get the so-called 'potential score'. It follows from our previous discussion that these actual and potential scores correct for input differences (in terms of *REG* and *EEO* instruction units), and avoid extreme specialization in *DUTCH* or *MATH* (through weight bounds). In addition, given that we focus on order-m inefficiency measures, it also accounts for possible outlier behavior. Finally, the use of conditional inefficiency measures corrects for environmental differences (in terms of *SES*, *B-MATH* and *B-DUTCH*).

Once we have derived distributions of actual and potential scores for the pupils in different school types, we investigate whether one school type is 'better' than another in a robust way, i.e., without assuming a specific parametric functional form to aggregate outcomes. To do so, we focus on First-order and Second-order Stochastic Dominance (FSD and SSD), two popular nonparametric dominance criteria in the risk and welfare literature. We start with FSD, and show how we can adjust it to correct for input and environmental differences between school types. Since we obtain inconclusive results, we next present SSD, which turns out to be a more powerful dominance criterion in the current setting.

5.1 Only efficiency matters: FSD

FSD can be characterized by the intuitive Pareto efficiency principle, which states that higher outcomes are always better. In words, one school type, say A, is better than another school type, say B, according to FSD, denoted $A \succsim_1 B$, if and only if welfare (denoted by W), measured by the average utility, is higher in A than in B for all increasing (differentiable) utility functions. Formally, for $\mathbb{U}_1 = \{U : \mathbb{R} \rightarrow \mathbb{R} \,|U' \geq 0\}$ the set of all increasing utility functions, we get

$$A \succsim_1 B \Leftrightarrow W_A - W_B = \int_0^\infty U dF_A - \int_0^\infty U dF_B \geq 0, \text{ for all } U \text{ in } \mathbb{U}_1,$$

with F_A and F_B the distribution functions of the actual scores for two school types. Using integration by parts, we obtain the following equivalent, implementable condition

$$A \succsim_1 B \Leftrightarrow F_A(y) - F_B(y) \leq 0, \text{ for all } y \in \mathbb{R}_+; \tag{1}$$

see, e.g., Lambert (2001). Notice that FSD is a robust ranking criterion, since it holds for all specifications of U within $\mathbb{U}_1$ (i.e. 'all utilitarians with increasing utility functions agree'). Still, it comes at a cost, since two distributions might turn out to be non-comparable.

Importantly, equation (1) does not take differences in inputs and school environment into account and would therefore be a rather blunt approach to assess school types. One way to correct for this, is to focus on the welfare difference between what is actually achieved (via the actual scores) and what could have been achieved (via the potential scores), i.e.,

$$\Delta W_{A|Z} = \int_0^\infty U dF_A - \int_0^\infty U dF_{A|Z}^{pot},$$

where $F_{A|Z}^{pot}$ is the distribution function of potential scores of the pupils in school type A conditional upon 'inputs' x and 'environment' z, collected in $Z = \{x;z\}$. Generally, higher values for $\Delta W_{A|Z}$ suggest better performance, as they indicate that —in aggregate welfare terms— the school type comes closer to potential achievement (while accounting for the given input and the environmental characteristics of the pupil population).

Given this, let $A \succsim_{1|Z} B$ denote that school type A is better than B according to FSD, corrected for inefficiency, measured conditionally upon Z; we get

$$A \succsim_{1|Z} B \Leftrightarrow \Delta W_{A|Z} - \Delta W_{B|Z} \geq 0, \text{ for all } U \text{ in } \mathbb{U}_1.$$

As before, using integration by parts, this equation can be equivalently expressed as

$$A \succsim_{1|Z} B \Leftrightarrow \left(F_A(y) - F^{pot}_{A|Z}(y)\right) - \left(F_B(y) - F^{pot}_{B|Z}(y)\right) \leq 0, \text{ for all } y \in \mathbb{R}_+. \quad (2)$$

Rewriting this equation, it consists of two components: a term $F_A(y) - F_B(y)$ which is the same as in equation (1), and a term $F^{pot}_{B|Z}(y) - F^{pot}_{A|Z}(y)$ which can be interpreted as the correction term. The following two simple examples illustrate the basic intuition. First, suppose that both school types are equally efficient, but have a very different pupil population in terms of z. In that case, the correction term $F^{pot}_{B|Z}(y) - F^{pot}_{A|Z}(y)$ tends to be equal to $F_B(y) - F_A(y)$ and will mitigate the first term $F_A(y) - F_B(y)$. Second, suppose both school types have the same inputs and the same environment for each pupil, but one school is better than the other in terms of actual scores, i.e. $F_A(y) - F_B(y) \leq 0$ everywhere. In this case, the correction term $F^{pot}_{B|Z}(y) - F^{pot}_{A|Z}(y)$ tends to zero and only the differences in the actual scores will play a role in assessing both school types.

Table 5 presents our results for the corrected FSD criterion in (2). We consider two extreme cases in terms of the specification of Z: the first case ($Z = \{REG{+}EEO;\emptyset\}$) does account for input differences but *not* for environmental differences (i.e., it is based on the unconditional inefficiency measure $\widetilde{\theta}^m_E$, which coincides with $\widetilde{\theta}_E(z_E)$ for z_E empty); the second case ($Z = \{REG{+}EEO;SES,B\text{-}MATH,B\text{-}DUTCH\}$) simultaneously takes account of input and *all* three environmental variables (i.e., it is based on the measure $\widetilde{\theta}_E(z_E)$, with z_E capturing *SES*, *B-MATH* and *B-DUTCH*). For each case, Table 5 reports the dominance relation between the row school type and the column school type: either the row school type 'dominates' or 'is dominated by' the column type, or the row type is not comparable to ('not comp. to') the column type. Two remarks are in order. First, following the usual practice, dominance is checked at 10 data points (equally spread over the common grid of both distributions), rather than at all points $y \in \mathbb{R}_+$. Second, we use a naive bootstrap procedure for statistical inference. That is, we calculate the proportion of the total number of bootstraps, i.e., 10000 drawings with replacement from the original sample, in which a certain result ('dominates', 'is dominated by', or 'not comp. to') was found.[6] In Table 5 we mention the most frequent result together with the corresponding 'naive' p-value, i.e., the proportion of times this result was found.

In terms of average test scores, we saw before that private schools outperform local public schools, while the latter in turn outperform Flemish public schools. Still, the results in Table 5, which pertain to the more robust FSD concept, do not allow us to conclude that one school type outperforms another in a significant way (i.e., using a naive p-value > 0.95,

[6] Notice that, from 5000 bootstrap samples onwards, the results remain stable.

which corresponds to a 5% significance level); in fact, this result holds for both (extreme) specifications of Z that we consider.

Table 5: Corrected FSD results.

$Z =$	$\{$REG+EEO;$\emptyset\}$		$\{$REG+EEO;SES,B-MATH,B-DUTCH$\}$	
	local public	Flemish public	local public	Flemish public
private catholic	dominates	not comp. to	dominates	not comp. to
	(0.5791)	(0.9966)	(0.7644)	(1.0000)
local public		not comp. to		not comp. to
		(1.0000)		(1.0000)

5.2 Equity also matters: SSD

We next include a preference for equality in addition to Pareto efficiency in our comparison of the aggregate performance of different school types. The SSD criterion simultaneously assesses efficiency and equity in a robust way. This dominance criterion can be characterized by the principle that higher outcomes are better (Pareto efficiency) and, additionally, the principle that more equal outcomes are better (Pigou-Dalton principle), i.e., more weight to lower scores. As a consequence, SSD is a necessary condition for FSD and leads to a more complete binary relation. According to SSD, school type A is better than school type B, denoted $A \succsim_2 B$, if and only if welfare (denoted by W, and again measured by the average utility) is higher in A than in B for all increasing and concave (twice differentiable) utility functions. Let $\mathbb{U}_2 = \{U : \mathbb{R} \to \mathbb{R}\,|U' \geq 0 \text{ \& } U'' \leq 0\}$ the set of increasing and concave utility functions; we get

$$A \succsim_2 B \Leftrightarrow W_A - W_B = \int_0^\infty U(x)\, dF_A(x) - \int_0^\infty U(x)\, dF_B(x) \geq 0, \text{ for all } U \text{ in } \mathbb{U}_2.$$

Twice integrating by parts leads to the equivalent condition

$$A \succsim_2 B \Leftrightarrow \int_0^y (F_A(x) - F_B(x))\, dx \leq 0, \text{ for all } y \in \mathbb{R}_+; \tag{3}$$

see again Lambert (2001).

Analogous to before, we propose a corrected version of the criterion in (3): $A \succsim_{2|Z} B$ means that school type A is better than B according to SSD, conditional upon inputs x and environment z, collected in $Z = \{x; z\}$. Formally, it is defined as

$$A \succsim_{2|Z} B \Leftrightarrow \Delta W_{A|Z} - \Delta W_{B|Z} \geq 0, \text{ for all } U \text{ in } \mathbb{U}_2;$$

and this dominance condition can be equivalently expressed as

$$A \succsim_{2|Z} B \Leftrightarrow \int_0^y \left[\left(F_A(x) - F_{A|Z}^{pot}(x)\right) - \left(F_B(x) - F_{B|Z}^{pot}(x)\right)\right] dx \leq 0, \text{ for all } y \in \mathbb{R}_+; \tag{4}$$

the interpretation is directly analogous to that of (2).

Table 6 presents our results. The interpretation of the different entries is similar to that of Table 5, but now pertains to the SSD criterion in (4). Interestingly, we now do find significant dominance relations, which is in sharp contrast to the FSD results in Table 5. If we consider the right column as the most fair comparison, then private schools significantly dominate public schools, while we cannot distinguish between the two types of public schools in a significant way. Note also that conditioning the efficiency scores plays a role when comparing both types of public schools.

Table 6: Corrected SSD results.

$Z =$	{REG+EEO;∅}		{REG+EEO;SES,B-MATH,B-DUTCH}	
	local public	Flemish public	local public	Flemish public
private catholic	dominates	dominates	dominates	dominates
	(0.9075)	(1.0000)	(0.9667)	(0.9998)
local public		dominates		not comp. to
		(1.0000)		(0.5014)

6 Conclusion

Focusing on educational efficiency, we have presented a nonparametric approach for analyzing public sector efficiency which also accounts for equity considerations. First, we have designed a nonparametric (DEA) model that is specially tailored for educational efficiency evaluation at the pupil level. It requires minimal *a priori* structure regarding the educational feasibility set and objectives. This is particularly convenient in the current context, which typically involves minimal *a priori* information. Next, we have argued that the First-order and Second-order stochastic dominance (FSD and SSD) criteria are particularly well-suited for comparing the educational efficiency of different school types; these nonparametric dominance criteria for comparing aggregate (school type) efficiency naturally complement our nonparametric model for evaluating individual (pupil level) efficiency. FSD is the appropriate criterion if only (Pareto) efficiency matters. By contrast, the more powerful SSD criterion is recommendable when (Pigou-Dalton) equity is important in addition to (Pareto) efficiency; such equity considerations are usually prevalent in the context of public sector efficiency evaluation. We have shown that our approach directly allows for adapting the methodology of Daraio and Simar (2005, 2007), to account for potential outlier behavior and environmental characteristics (*in casu* the pupils' educational environment) in the efficiency assessment. Although our application concentrates on educational efficiency, we believe that the presented approach is also more generally useful for efficiency evaluation in the public sector: such evaluation often (1) involves little *a priori* information about the underlying feasibility set and objectives, and (2) focuses on comparisons of the aggregate efficiency of different groups, in which (3) equity considerations are important.

Our application demonstrates the practical usefulness of our approach. First, we have investigated the impact of the 'environmental characteristics' socio-economic status (*SES*), and begin-level in mathematics (*B-MATH*) and language proficiency (*B-DUTCH*) on the educational efficiency for individual pupils. In line with our prior expectations, we find that all three environmental variables have a positive first-order effect on educational efficiency: on average, higher *SES*, *B-MATH* or *B-DUTCH* values systematically entail higher educational efficiency for individual pupils; although for high *SES* values, the first-order effect is negative. In addition, we find that the (average) second order effects are always negative and more pronounced for low values of the environmental variables, which suggests that the positive first-order effect prevails in particular when the initial *SES*, *B-MATH* and *B-DUTCH* status is low. Although we find stronger effects for *B-MATH* and *B-DUTCH* than for *SES*, we believe that our results convincingly support that all three environmental variables should simultaneously be accounted for to obtain a fair efficiency evaluation. Next, we have compared the aggregate efficiency of private schools, local public schools and Flemish public schools. Focusing on FSD, we find that no school type robustly dominates another school type; we conclude 'non-comparability' in all pairwise comparisons. However, the story changes dramatically if we focus on SSD. When accounting for the diverging environmental characteristics of the pupil populations, we find that private schools significantly dominate both types of public schools. In addition, our results suggest that local public schools outperform Flemish public schools, but this result is not supported in a significant way. These results are in line with the seminal work of Coleman *et al.* (1982) and are consistent with the mainstream literature. Given that our aggregate efficiency comparisons account for both equity and environment, we consider them as most fair in the (public sector) evaluation context under study.

References

[1] Altonji, J.G., T.E. Elder and C.R. Taber (2005), Selection on observed and unobserved variables: assessing the effectiveness of catholic schools, *Journal of Political Economy* 113 (1), 151-184.

[2] Cazals, C., J. Florens and L. Simar (2002), Nonparametric frontier estimation: a robust approach, *Journal of Econometrics* 106, 1–25.

[3] Cain, G.G., A.S. Goldberger (1983), Public and private schools revisited, *Sociology of Education* 56 (4), 208-218.

[4] Chambers, R., R. Färe and Y. Chung (1998), Profit, directional distance functions, and nerlovian efficiency, *Journal of Optimization Theory and Applications* 98, 351-364.

[5] Coleman, J., T. Hoffer and S. Kilgore (1982), Cognitive Outcomes in Public and Private Schools, *Sociology of Education* 55 (2), 65-76.

[6] Daraio, C. and L. Simar (2005), Introducing environmental variables in nonparametric frontier models: a probabilistic approach, *Journal of Productivity Analysis* 24, 93–121.

[7] Daraio, C. and L. Simar (2007), *Advanced robust and nonparametric methods in efficiency analysis. Methodology and applications*, Series: Studies in Productivity and Efficiency, Springer.

[8] Deprins, D., L. Simar and H. Tulkens (1984), Measuring labor efficiency in post offices, *The Performance of Public Enterprises: Concepts and Measurements*, M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

[9] Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1997), Budget-constrained frontier measures of fiscal equality and efficiency in schooling, *Review of Economics and Statistics* 79, 116-124.

[10] Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1999), Anticipating the consequences of school reform: a new use of DEA, *Management Science* 45, 608-620.

[11] Lambert, P.J. (2001) *The Distribution and Redistribution of Income. A Mathematical Analysis*, 3rd edition, Manchester: Manchester University Press.

[12] Levy, H. (1992) Stochastic dominance and expected utility: survey and analysis, *Management Science* 38(4), 555-593.

[13] Mardia, K.V., J.T. Kent and J.M. Bibby (1979), Multivariate analysis, Academic Press New York

[14] Portela, M. and E. Thanassoulis (2001), Decomposing school and school type efficiencies, *European Journal of Operational Research* 132, 357-373.

[15] Reynders, T., I. Nicaise and J. Van Damme (2005), Longitudinaal onderzoek in het basisonderwijs. De constructie van een SES-variabele voor het SIBO-onderzoek, *LOA-rapport* 31.

[16] Simar, L. (2003), Detecting outliers in frontier models: a simple approach. *Journal of Productivity Analysis* 20, 391–424.

[17] Tulkens, H. (1993), On FDH efficiency analysis: some methodological issues and applications to retail banking, courts and urban transit, *Journal of Productivity Analysis* 4 (1/2), 183-210.

[18] Varian, H.R. (1984), The nonparametric approach to production analysis, *Econometrica* 52, 579-598.

[19] Worthington, A.C. (2001), An empirical survey of frontier efficiency measurement techniques in education, *Education Economics* 9, 245-268.